Le travail de l'Intelligence Artificielle : concevoir et entraîner un outil de pseudonymisation automatique à la Cour de Cassation

Camille Girard-Chanudet

Le travail de l'Intelligence Artificielle : Concevoir et entrainer un outil de pseudonymisation automatique à la Cour de Cassation[1]

Camille Girard-Chanudet. 2023. « Le travail de l'Intelligence Artificielle : Concevoir et entrainer un outil de pseudonymisation automatique à la Cour de Cassation », RESET, n°12.

Responsable de la mise en œuvre de l'*open data* des décisions de justice, la Cour de Cassation développe, depuis 2018, un algorithme d'intelligence artificielle de pseudonymisation automatique de la jurisprudence. À partir d'une enquête ethnographique, cet article rend compte des modalités spécifiques d'articulation du travail des technicien·nes, annotateur·trices et juristes impliqués dans la mise en œuvre de ce projet. L'analyse des activités de catégorisation sous-tendant le fonctionnement de l'IA, alliant logiques théoriques et empiriques, met en évidence l'importance des choix socialement situés qui contribuent à donner corps à des outils pourtant considérés comme autonomes, neutres et objectifs.

[1] Recherche financée par le programme paris région phd[2]

Introduction

Depuis le début des années 2010, les techniques dites d' « intelligence artificielle » (IA), basées sur le traitement algorithmique de grandes quantités de données numériques, investissent un nombre croissant de secteurs d'activité, de la finance à l'automobile en passant par le marketing ou la création artistique. Les services publics, y compris régaliens, n'échappent pas à cette dynamique. C'est notamment le cas de la justice française, qui fait l'objet d'expérimentations dans le domaine de l'IA depuis l'adoption du principe mise à disposition publique de la jurisprudence par la loi pour une République numérique en 2016. Développés par des *start-ups* de la *legal tech* nouvellement formées, par des éditeurs juridiques traditionnels et, à partir de 2018, par le Ministère de la justice et les cours suprêmes, ces outils dits « prédictifs » ont vocation à traiter ce flux inédit de 3 millions de décisions annuelles. Par le biais d'une analyse massive des textes des décisions, il s'agit ainsi d'inférer à partir de situations passées des solutions probables pour des cas inédits – que ce soit pour « prédire » le montant d'indemnisation d'un préjudice corporel (outil Datajust du ministère de la justice), pour estimer l'issue d'un contentieux en fonction de la juridiction auquel il est présenté (Jurisdata Analytics) ou encore pour anonymiser les décisions avant leur publication (Judilibre à la Cour de Cassation). Les premiers usages – controversés et aux résultats mitigés – de ce type d'outils au sein des juridictions ont déjà fait l'objet de quelques enquêtes axées vers leur prise en main, notamment en France (Lacour et Piana, 2019 ; Licoppe et Dumoulin, 2019) et aux États-Unis (Christin, Rosenblat et Boyd, 2015)

Souvent qualifiés de « boites noires » (Cardon, 2015 ; Eubanks, 2018 ; Pasquale, 2016), ces outils sont fréquemment associés à des idéaux d'immatérialité, de neutralité et d'objectivité qui seraient garantis par leur éloignement des biais inhérents aux processus de décision humains (Boyd, 2016 ; Crawford, 2021). Dans le domaine judiciaire, ils sont notamment supposés compenser la possible inconstance de processus de jugement humains soumis aux contingences émotionnelles et temporelles (Danziger, Levav et Avnaim-Pesso, 2011). Pourtant, loin de cette aura « magique » (Elish, 2017 ; Elish et Boyd, 2018), les techniques contemporaines d'IA sont le fruit du travail de multiples équipes de recherche, d'investissements conséquents et d'évolutions théoriques profondes, et ce depuis les années 1950 (Cardon, Cointet et Mazières, 2018). Par ailleurs, leur développement et leur mise en œuvre au sein d'organisations spécifiques sont conditionnés à la mobilisation d'équipes dédiées, constituées d'ingénieur·es informatiques mais également

de « petites mains » annotatrices, auxquels plusieurs travaux ont été consacrés (Casilli, 2019 ; Roberts, 2020). L'ensemble de ces professionnel·les oriente en coulisses (Goffman, 1973), par leurs activités de traitement de l'information et d'optimisation algorithmique, le fonctionnement et les résultats que produit l'IA, sans que ces choix ne soient nécessairement rendus visibles au moment de la mise en service de ces outils. Le travail entourant la conception d'outils d'IA s'apparente en ce sens à une manifestation du « travail invisible des données » auquel une série de travaux fondateurs a été consacrée (notamment Denis, 2018 ; Denis et Pontille, 2012 ; Edwards et al., 2011 ; Goëta, 2016).

Cet article vise à mettre en évidence l'importance de ces agencements organisationnels et des activités professionnelles entourant le développement d'outils d'IA par le biais d'une étude de cas : celui de la conception d'un outil de pseudonymisation des décisions par apprentissage automatique (*machine learning*) au sein de la Cour de Cassation – visant à occulter les principaux termes identifiants des personnes physiques du texte des décisions.[2] Positionnée très tôt comme responsable de la mise en œuvre de l'*open data* jurisprudentiel, la juridiction suprême s'est dotée, depuis 2019, d'une équipe dédiée à la conception et au suivi d'un moteur d'IA chargé de la pseudonymisation des décisions préalable à leur diffusion publique. Cette tâche, essentielle au respect du cadre réglementaire européen de protection de la vie privée, ne pourrait en effet être réalisée manuellement en raison du volume de décisions concernées. Contrairement à un moteur « par règles » exécutant systématiquement certaines consignes (par exemple, supprimer automatiquement le mot figurant après « Madame » dans une décision), un moteur par apprentissage automatique n'applique pas des principes définis, mais « apprend » à partir d'une grande quantité de cas particuliers. Son fonctionnement dit « prédictif » est statistique : chaque terme contenu dans une décision est « vectorialisé » (représenté par des chiffres) et positionné dans un espace à *n* dimensions au sein duquel figurent les catégories idéal-typiques des entités à identifier (nom, prénom, adresse…). La proximité de chacun des termes ainsi positionnés dans l'espace avec les éléments recherchés permet de conclure avec un certain degré de certitude à la concordance entre les mots présents dans la décision et la catégorie à pseudonymiser (si un terme est « suffisamment » proche du pôle « Nom » - 96% de concordance par exemple – on pourra en déduire qu'il s'agit effectivement d'un nom et l'occulter automatiquement). Pour réaliser ce traitement statistique sur des décisions inconnues, le moteur d'apprentissage automatique est préalablement entrainé

[2] La pseudonymisation se distingue de l'anonymisation en ce qu'elle ne suppose pas d'impossibilité définitive de réidentification des individus concernés (qui peuvent être identifiés par croisement d'informations notamment).

sur des jeux de plusieurs centaines de décisions annotées manuellement, c'est-à-dire sur lesquelles les mots correspondant aux catégories prédéterminées ont été identifiés.

Ma recherche s'appuie sur une ethnographie menée entre janvier et juin 2021 au sein du pôle open data du service de la documentation, des études et du rapport (SDER) de la Cour de Cassation, chargé du développement de l'outil d'IA ainsi que de l'interface d'annotation des décisions. Il fait également appel à des entretiens semi-directifs conduits avec l'ensemble des membres de l'équipe (n= 20), ainsi qu'à l'analyse de corpus documentaires relatifs aux infrastructures techniques et à l'organisation interne du travail.

Il s'agit ainsi, dans une première partie, d'entrer dans l'épaisseur de l'agencement professionnel nécessaire au développement de cet outil d'IA, dans la lignée des premières ethnographies du travail de construction algorithmique (Forsythe, 2001 ; Jaton, 2019). Je mettrai ainsi en évidence la multiplicité des tâches sous-tendant le cadrage de l'outil et son fonctionnement, ainsi que l'importance du travail d'articulation nécessaire à la rencontre de groupes professionnels parfois éloignés dans le cadre de de ce projet (Star et Strauss, 1999 ; Strauss, 1985). L'étude des opérations de catégorisation des entités à pseudonymiser, oscillant entre définition théorique et ajustements empiriques (Bowker et Star, 1999), et conduites à différents niveaux par l'ensemble de l'équipe, permettra dans une seconde partie de montrer que la conception d'un outil d'IA est indissociable de choix socialement situés, contribuant par le travail sur les données à donner corps à l'algorithme et aux résultats que celui-ci produit.

Le pôle open data de la Cour de Cassation : un écosystème récent, pluriel et collaboratif

La construction d'un outil d'IA suppose la mobilisation de savoir-faire multiples, dont l'articulation dans un contexte donné conditionne en partie le fonctionnement de l'outil technique en production, en même temps qu'il éclaire sur les objectifs associés à son déploiement. Si de multiples configurations professionnelles peuvent sous-tendre la conception d'instruments de ce type, la forme que prend cet assemblage d'expertises au sein de la Cour de Cassation est particulièrement rare. Le pôle *open data*, en charge de la conception d'un outil de pseudonymisation automatique des décisions de justice, réunit en effet en son sein, dans une même aile du Palais de Justice de l'Ile de la Cité, l'ensemble des profils nécessaires au développement d'un outil

d'IA – là où, pour ce type de projets, tout ou partie des tâches sont fréquemment externalisées (Roberts, 2020).

La constitution de l'équipe, loin d'être évidente, s'effectue au gré de l'évolution des besoins, de la prise d'importance du projet et des contraintes associés aux opérations de recrutements. Le pôle open data fait ainsi collaborer un groupe de profils techniques responsables de l'infrastructure logicielle (1.1) et une équipe d'agent·es administratif·ves de catégorie C chargée de l'annotation des décisions en vue de l'entrainement et de la correction de la machine (1.2), dont les missions s'articulent étroitement à l'expertise juridique des magistrat·es de la Cour (1.3). L'ensemble fonctionne suivant une logique projet, corrélée à l'important soutien de l'écosystème de la « transformation de l'action publique » à ce chantier d'ouverture de de données publiques (1.4).

La composition d'une équipe technique : définition des contours et mise en oeuvre du projet open data

En 2016, lors de l'adoption de la loi pour une République Numérique introduisant l'*open data* des décisions judiciaires, la Cour de Cassation était loin de disposer des compétences nécessaires à l'exercice de la mission qui allait lui être confiée. Son service informatique, chargé principalement de l'administration des applicatifs métiers de la Cour et aux effectifs limités, ne comportait aucun·e spécialiste d'*open data* ni d'IA.

La Cour de Cassation décide en 2018 de se doter de compétences afin d'assurer directement la pseudonymisation des décisions de justice. Cette stratégie, portée par les échelons hiérarchiques les plus élevés de la Cour, a en partie orienté les processus d'ouverture des données. La composition évolutive de l'équipe technique est en effet indissociable de la définition des objectifs stratégiques et opérationnels du projet, à laquelle elle participe au titre de son expertise, ainsi que de la mise en place d'une infrastructure matérielle dédiée que son activité a progressivement nécessité.

Le recrutement de l'équipe technique (qui sera encore amenée à évoluer) s'est effectué en plusieurs temps. En tout, ce sont trois développeurs, deux *data scientists* et un designeur qui rejoignent la cour entre 2019 et 2021, avec pour missions successives la construction d'un moteur d'apprentissage automatique de pseudonymisation des décisions ainsi que la conception d'une interface d'annotation. Les tâches accomplies au quotidien par cette équipe technique sont multiples. Les *data scientists* travaillent de façon relativement autonome à l'entretien et l'amélioration du moteur de pseudonymisaton automatique ; il s'agit de mettre en place de nouvelles catégories en fonction des décisions prises par l'équipe, de tester des modèles d'apprentissage « à l'état de l'art » -

développés par les grandes entreprises du web et diffusés en *open source* - et de trouver les meilleures combinaisons de paramètres afin d'optimiser les calculs. L'équipe de développement et de design travaille en parallèle à la conception de l'interface d'annotation ; elle identifie, en lien avec les agents d'annotation, les fonctionnalités attendues du logiciel, réalise les opérations de codage nécessaires à leur réalisation et complexifie progressivement sa structure.

Les membres de l'équipe technique se placent en décalage des professionnel·les évoluant habituellement au sein de la Cour de Cassation (magistrat·es, greffier·es, professions juridiques de façon générale), tant par leurs missions que par leurs parcours professionnels. Il s'agit en effet de jeunes recrues, essentiellement masculines et issues pour la plupart du secteur privé, dont l'expérience professionnelle est axée vers l'opérationnalité et l'efficience – bien que leur intérêt pour le service public ait constitué un facteur déterminant dans leur recrutement. Cette équipe, parce qu'elle est la seule dépositaire des compétences techniques nécessaires au bon fonctionnement du projet, joue un rôle essentiel dans sa propre structuration ainsi que dans la définition de ses objectifs de travail. Les premiers membres de l'équipe ont ainsi participé activement au recrutement des nouveaux·elles venu·es, et c'est en partie sur la base de leur retour d'expérience qu'a été développé, à la suite de la mise en service du moteur d'apprentissage automatique, le projet d'ampleur de conception d'une interface destinée à faciliter et à optimiser le travail de l'équipe d'annotation.

La mise en place de l'*open data* a donc conduit à la structuration d'une équipe particulière au sein de la Cour de Cassation, dont les activités sont étroitement entremêlées avec celles des profils juridiques – décisionnaires concernant les grandes orientations du projet – mais disposant de marges de manœuvre et de capacités de cadrage importantes pour la réalisation de ces missions en raison de leur possession exclusive des compétences techniques adaptées.

Les missions accomplies par l'équipe *open data* amènent par ailleurs la Cour de Cassation à se doter d'une infrastructure technique inédite par rapport à ses activités traditionnelles – et fonctionnant parfois de façon autonome par rapport au système informatique de l'institution. C'est par exemple le cas de la tour GPU, serveur de calcul nécessaire à l'entrainement des algorithmes d'apprentissage automatique, acquis spécifiquement à cette fin en 2019 à la demande de l'équipe technique. Contrairement aux autres serveurs de la Cour, celui-ci n'est pas géré par le service informatique mais directement par les *data scientists*. Dans la lignée de travaux menés sur la matérialité de l'IA et ses impacts organisationnels (Crawford & Joler, 2018), cette tour de calcul

matérialise à la fois les évolutions infrastructurelles induites à la Cour de Cassation par la mise en œuvre de ce chantier d'IA, et la relative autonomie matérielle et opérationnelle octroyée à l'équipe technique spécialisée en charge de ce projet.

L'équipe d'annotation : petites mains et rouage essentiel de la pseudonymisation

Parallèlement à la constitution de l'équipe technique, la Cour de Cassation s'est dotée d'une équipe d'annotation dont l'effectif s'est étoffé jusqu'à atteindre une quinzaine de personnes au moment de mon enquête. Cette équipe réalise une part essentielle du « travail des données » (Denis, 2018) nécessaire au bon fonctionnement d'un outil d'IA : elle vérifie et corrige à la main l'annotation des décisions de justice afin d'entrainer, puis de corriger, le moteur de pseudonymisation automatique. Concrètement, il s'agit pour ces travailleur·ses de parcourir une par une des décisions de justice pré-annotées par la machine – qui indique pour chaque groupe de mots s'il s'agit d'un terme considéré comme « identifiant » (nom, adresse, date de naissance, adresse mail…) –, de vérifier la pertinence de cette annotation et, le cas échéant, de la corriger. De telles opérations sont indispensables non seulement à l'amélioration des performances du moteur automatique, qui « apprend » à reconnaître des entités inédites partir de jeux de données correctement annotés (les « données *gold* »), mais également au respect de ses obligations légales par la Cour de Cassation, qui, en tant que responsable de traitement, ne doit pas diffuser des décisions à la pseudonymisation inexacte, en particulier s'agissant de certains types de contentieux sensibles.

L'existence d'une équipe responsable d'une telle mission au sein de la Cour de Cassation est une spécificité dans le paysage de l'IA : ce type de travail étant fréquemment laissé aux *data scientists* (avec un rendement limité) ou sous-traité à des plateformes spécialisées. Ces dernières ont généralement recours à des travailleur·ses « indépendant·es » réalisant à la chaine des micro-tâches, souvent mal rémunérées, sur le modèle d'*Amazon Mechanical Turk* (Casilli, 2019 ; Roberts, 2020). L'équipe de la Cour de Cassation est au contraire constituée de fonctionnaires, agent·es administratif·ves de catégorie C, pour la plupart des femmes, issu.es du monde de la justice ou en détachement, et d'agents contractuels. Cette configuration s'explique à la fois par le caractère confidentiel des données à traiter, difficilement exportables, et par le contrôle direct que l'institution se donne ainsi les moyens d'exercer sur l'activités de ces agents. L'équipe est en effet placée sous l'autorité d'une directrice des services de greffe judiciaires, chargée de l'articulation de l'activité des agents entre eux et avec le reste du pôle *open data* - à la périphérie duquel ils demeurent.

Si les conditions d'emploi de l'équipe d'annotation de la Cour de Cassation sont différentes de celles de leurs homologues, « *crowdworkers* », réalisant le même type de tâches pour le compte de plateformes multinationales, le travail n'en demeure pas moins fastidieux et répétitif (Lagerie & Santos, 2018 ; Roberts, 2020). Les agent·es chargés de cette mission enchainent le traitement minutieux de plusieurs dizaines de décisions quotidiennes sur l'écran de leur ordinateur, tout en étant conscient·es des potentiels risques en termes de protection de la vie privée pour les justiciables en cas d'erreur de leur part. Iels mettent de ce fait en place diverses stratégies afin de maintenir leur niveau d'attention stable (pauses régulières, exercices de respiration et d'étirements – parfois conseillés par des professionnel·les de la santé –, écoute de musique, réglage de la luminosité de l'écran…).

De telles équipes de « petites mains » (Denis & Pontille, 2012) sont indissociables du développement de ce type d'outils techniques : constituer des jeux d'apprentissage fiables et corriger la machine (notamment en vue de la publication d'informations potentiellement sensibles) demande de l'attention et du temps. Parce que les décisions, issues d'un monde réel changeant et imprévisible, sont toujours susceptibles de contenir des types d'éléments identifiants non anticipés, les activités automatisées de l'IA reposent sur un travail continu d'annotation et de vérification, dont la trace est invisibilisée dans les résultats produits par la machine.

C'est dans l'objectif de facilitation et d'optimisation du travail de cette équipe (limitée en taille et en vue de la gestion à terme d'un flux d'environ 3 millions de décisions annuelles rendues publiquement) qu'a été monté le projet de conception d'une interface numérique intuitive et efficace d'annotation des décisions. Associé à une priorisation des circuits de relecture des décisions (il s'agit de définir quels types de décisions doivent être impérativement relues par l'équipe d'annotation, et quels types de décisions, parce que moins sensibles ou parce que le moteur automatique les traite particulièrement bien, pourraient être publiées sans relecture humaine), ce projet a pour objectif de conduire à terme à une augmentation du rendement de la pseudonymisation. Cet objectif d'efficience, posé dès l'origine par les initiateurs du projet, concentre donc deux des enjeux majeurs du développement de l'IA : la nécessité, d'une part, d'amélioration de la fiabilité des résultats du moteur d'apprentissage automatique et l'optimisation, d'autre part, du temps de traitement humain de la part incompressible de décisions à annoter manuellement.

Initié en septembre 2020 avec le recrutement de deux développeurs et un designeur, la conception de cette nouvelle interface d'annotation constitue une charnière centrale entre le travail des équipes techniques et d'annotation au sein du pôle *open data*. Généralement isolé·es du reste du pôle, les agents d'annotation ont été amenés à échanger régulièrement avec les technicien·nes dans tout au long de ce projet. Eloigné·es des activités techniques auxquels iels contribuent pourtant pleinement (plusieur·es m'ont indiqué en entretien ne « rien connaitre à l'IA »), iels se sont ainsi retrouvés (parfois avec surprise) au cœur du travail de conception d'une application d'annotation. En participant à des tests d'utilisation et en signalant les *bugs* et difficultés lors de la mise en service de l'interface, l'équipe d'annotation a joué un rôle essentiel dans son développement. Utilisatrice active d'un produit « beta » en évolution constante, elle a ainsi réalisé une part significative du travail itératif de conception de l'interface (Neff et Stark, 2004).

Cette implication de l'équipe d'annotation dans les processus techniques a eu un effet ambivalent sur la position de ces agents au sein du pôle *open data*. Si elle a induit, d'une part, un renforcement de l'engagement d'une partie de l'équipe sentant son expertise reconnue et prise en compte, elle a également mis en évidence la position annexe de ces travailleur·ses par rapport aux autres membres du pôle. Malgré leur apport essentiel au développement de l'interface, les agents de l'équipe d'annotation ne participent pas aux réunions hebdomadaires de l'équipe projet, sont peu informé·es des enjeux et échéances sous-tendant le projet, et pas impliqué·es dans les processus décisionnels le concernant. Iels apparaissent ainsi davantage comme un public-test facilement accessible que comme de véritables collaborateur·trices

Un pôle open data coordonné par l'expertise juridique

Les activités des équipes techniques et d'annotation du pôle *open data* de la Cour de Cassation sont enfin étroitement articulées au travail juridique, socle hiérarchisé et codifié de l'institution. L' « expertise-métier » sous-tend et encadre l'ensemble des réalisations du pôle *open data*, à la fois en termes d'orientations stratégiques (et souvent diplomatiques, en lien avec les acteurs extérieurs au service) et s'agissant des fondements conceptuels et juridiques du projet de pseudonymisation. Elle dispose d'un poids prépondérant dans la définition de l' « arc de travail » du pôle *open data* (Strauss, 1985), c'est-à-dire dans le cadrage des objectifs du projet et des tâches associées à leur réalisation. En principe (la pratique est moins définie), les activités techniques et d'annotation sont ainsi supposées exécuter des orientations décidées par des magistrat·es aux différents échelons hiérarchiques traditionnels de la Cour de Cassation – la validation des grandes lignes du projet *open data* dépendant de la validation de la Première Présidence. Garante de cet équilibre

institutionnel et juridique, et du respect des lignes directrices déterminées par des groupes de travail constitués à ce sujet, une conseillère référendaire fortement qualifiée en gestion de projet est ainsi chargée de la coordination de l'ensemble de l'équipe. Une auditrice la seconde dans le cadrage juridique du projet, guidé par une nécessité d'arbitrage entre impératif de publication (et de lisibilité) des décisions, et respect de la vie privée des personnes physiques impliquées. Concrètement, ce travail juridique implique d'une part la définition des types d'entités à pseudonymiser (nom, prénom, adresse, n° INSEE…), et d'autre part la détermination de leur activation (ou non) en fonction des types de contentieux (on occulte davantage d'informations dans des décisions concernant le droit de la famille que dans des décisions traitant de droit des sociétés, par exemple). Ces tâches irriguent les activités de l'ensemble du pôle *open data*, des *data scientists* configurant leurs modèles sur la base de ces orientations aux annotateur·trices suivant les préconisations juridiques, en passant par les développeurs mettant en place les catégories au sein de l'interface. Novices dans un domaine du numérique auquel elles n'avaient pas été formées préalablement, les magistrates du pôle *open data* se sont familiarisées avec des enjeux et termes techniques avec lesquels elles travaillent désormais au quotidien, de la même façon que les profils techniques se familiarisent avec la matière juridique, dans une forme d'hybridation des savoirs juridico-informatiques.

Une articulation efficace basée sur une logique projet, inscrite dans le cadre de la « transformation de l'action publique »

Le pôle *open data* de la Cour de Cassation réunit donc en son sein, autour du projet de pseudonymisation et de mise à disposition des données de jurisprudence, une équipe professionnelle hétérogène constituée d'une dizaine de personnes (profils techniques et juridiques) auxquels s'ajoute la quinzaine d'agent·es d'annotation et leur cheffe d'équipe. La rencontre de ces expertises et cultures professionnelles parfois très éloignées ne génère pas de tensions apparentes au premier regard au sein de l'environnement de travail, davantage caractérisé par la coopération que par la conflictualité. De fait, la répartition des responsabilités semble claire et déterminée en fonction du domaine d'expertise de chacun·e ; cette structuration s'accompagne de l'omniprésence d'opérations de traduction (Callon, 1986) entre les différents groupes d'expertise, aboutissant à l'établissement de bases communes de compréhension des problématiques en vue de leur résolution. L'équipe technique consacre ainsi une partie conséquente de son temps à l'explicitation de ses activités (parfois à l'aide de supports visuels) de façon à rendre accessible les enjeux techniques au reste du groupe, et à faire participer les

« non-technicien·nes » à certains arbitrages. Réciproquement, le vocabulaire juridique est régulièrement converti en termes compréhensibles par tout·es : circulant en régulièrement entre des registres professionnels distincts, les membres du pôle *open data* construisent par leurs efforts de traduction des « points de passage » nécessaires à une coopération efficace (Star et Griesemer, 1989)

Si l'analyse à un degré plus fin de la conduite conjointe de certaines opérations par l'ensemble de l'équipe doit conduire à nuancer ce constat de fluidité, on peut toutefois tenter d'expliquer ce fonctionnement global.

Il faut pour ce faire resituer l'élaboration du projet *open data* dans un cadre institutionnel plus large. La mise en place d'un projet d'IA visant à automatiser la pseudonymisation des décisions de justice en vue de leur publication a inscrit pleinement, et ce dès l'origine, la Cour de Cassation dans le paysage de la « transformation de l'action publique » (Alauzen, 2019a ; Bezes, 2009) et plus précisément des « réformes » numériques de la justice (Vauchez & Willemez, 2007). Programme interministériel porté par la Direction Interministérielle de la Transformation Publique et Etalab (service de la Direction interministérielle du numérique en charge d' « ouvrir, de partager et de valoriser les données publiques »), cette initiative vise à encourager et à soutenir la modernisation et la numérisation des administrations par le biais de programmes. L'un d'eux, « Entrepreneurs d'intérêt général » (EIG), prévoit le financement par Etalab de trois postes techniques, ayant vocation à intégrer des administrations porteuses d'un projet de « transformation » pour une durée de 10 mois. C'est par l'intermédiaire de ce programme que deux *data scientists* et un développeur ont rejoint la Cour de Cassation en 2019 pour la conception du moteur d'apprentissage automatique et du moteur de recherche de jurisprudence (avant que ces trois postes ne soient pérennisés), suivis par deux développeurs et un designeur en 2020 pour la conception d'une interface d'annotation.

Le recrutement particulier de ces profils techniques est susceptible d'expliquer, au moins en partie, le fonctionnement de l'équipe *open data*. Les « entrepreneurs d'intérêt général », rémunérés par Etalab et par le ministère de la Justice, et intégrés au large réseau de ce programme, sont sélectionnés pour répondre à un « défi » dont les objectifs ont été déterminés préalablement à leur arrivée. Ils disposent d'une autonomie relative, qu'iels apprécient et défendent particulièrement. L'un des EIG explique ainsi en entretien :

« L'avantage des EIG, c'est qu'on n'a pas « officiellement » de chef en fait. Ici on travaille sous la houlette et avec les conseils [des profils juridiques], mais en

soi on n'a pas de patron, donc si on veut faire quelque chose on le fait, [les profils juridiques] ne peuvent pas faire grand-chose contre nous. Bon, il faut un esprit d'équipe bien sûr, je veux pas dire qu'on fait n'importe quoi, mais en soi on n'a pas de lien hiérarchique direct. Et vu que c'est nous qui avons le savoir-faire, ben ça marche bien »

L'intégration du chantier *open data* au programme EIG a ainsi contribué à la structuration du pôle suivant une logique « projet » (Boltanski & Chiapello, 1999), avec des objectifs finaux (un « cahier des charges »), une attribution des rôles et un calendrier très précis. Ce mode de fonctionnement se répercute au quotidien sur l'organisation de l'équipe, qui fait un usage très important d'outils de gestion, de priorisation et d'attribution de ses activités (rétro-plannings…). Une telle structuration oriente l'équipe vers l'opérationnalité (qui se manifeste par l'objectif premier de mise en service d'un *minimum viable product* - version fonctionnelle a minima des outils, qui peuvent ensuite être améliorés) et réduit les risques de friction autour d'enjeux de fond ou de long terme au sujet desquels les technicien·nes, voué.es à rester un temps limité dans l'institution, n'ont concrètement que peu à se préoccuper (cela ne les empêchant pas toutefois d'exprimer des avis et suggestions).

À cela s'ajoute la constitution progressive et par consensus de l'équipe, chaque nouveau recrutement ayant été réalisé en fonction de besoins déterminés, et chaque nouveau membre étant choisi·e non seulement pour ses compétences, mais également pour son adéquation avec les valeurs portées par l'institution pour ses capacités à travailler au sein d'une équipe pluridisciplinaire.

La conception et la gestion d'un outil de pseudonymisation par IA nécessite donc la collaboration d'une équipe aux compétences plurielles. Loin de l'image d'autonomie généralement associée à ce type de dispositifs techniques, la charge de travail humain induite dans le fonctionnement d'un tel outil est conséquente : celui-ci mobilise au quotidien plus de 20 personnes à la Cour de Cassation. La coordination de ces profils (juridiques, techniques, d'annotation) implique la mise en place explicite et implicite de processus organisationnels et décisionnels spécifiques, ainsi que l'établissement d'une infrastructure technique dont la forme est liée à contraintes contextuelles. C'est au sein de cet agencement sociotechnique spécifique que prend corps l'outil de pseudonymisation automatique : les objectifs qui lui sont assignés et les processus déterminés pour y répondre (en termes de fonctionnement algorithmique comme de travail humain) sont directement liés à la configuration au sein de laquelle il est conçu. L'analyse du travail de catégorisation des entités à pseudonymiser, distribué entre les différentes parties prenantes au

projet et oscillant entre logiques déductives et inductives, permettra de mieux comprendre cette idée.

Distribution et circulation du travail de catégorisation

Le membres du pôle *open data* accomplissent au quotidien des tâches d'ordre divers (codage, revue de code, revue des algorithmes à l'état de l'art, conduite de tests, relecture, annotation, correction, évaluations statistiques…). Parmi ces opérations, l'une d'elles tient une place particulière, parce que sa réalisation protéiforme implique l'ensemble des membres de l'équipe, et qu'elle irrigue l'architecture de la pseudonymisation. Il s'agit du travail de catégorisation des entités à pseudonymiser. Tâche structurante des organisations modernes, souvent invisibilisée et profondément sociale (Desrosières & Thévenot, 2002 ; Gardey, 2008), la catégorisation représente également un élément essentiel pour tout projet d'apprentissage automatique supervisé, l'objectif principal assigné à la machine étant de reconnaître dans les données qui lui sont proposées les entités correspondant à des catégories prédéterminées. Dans le cas du moteur de pseudonymisation, ces catégories correspondent aux grands types d'informations ayant vocation à être occultés. Ceux-ci représentent une liste d'une quinzaine de catégories environ, dont le scope dépasse les simples informations identitaires telles que le nom et le prénom (on y retrouve par exemple les adresses mail, comptes bancaires, dates de décès, cadastre…). Le but assigné au moteur d'IA est ainsi d'identifier les termes correspondant à ces catégories dans les décisions de justice lui étant soumises, de les labelliser de façon adéquate, et, enfin, de procéder à la pseudonymisation, c'est-à-dire au remplacement de termes concernés (Madame Dupont devenant ainsi Madame [X], par exemple). Activité principale du moteur d'IA, la catégorisation (entendue à la fois comme définition et attribution des catégories) occupe également une part conséquente du temps de l'équipe ; elle fait l'objet de débats en termes de faisabilité technique comme de pertinence logique et juridique, de doutes et d'enquêtes conduisant à reconfigurer de façon continue les principes guidant le moteur de pseudonymisation. La mise en œuvre de la catégorisation est révélatrice de l'importance des choix collectifs et individuels dans la détermination du fonctionnement de l'IA en construction – éclipsés, une fois mise en service, par son apparente autonomie. Ceux-ci conditionnent les résultats auxquels la machine aboutit, mettant à mal l'idée d'une objectivité purement mathématique mise en avant dans nombre des discours publics entourant l'IA (Ganascia, 2017).

Croisement de logiques déductive et inductive : la distribution hiérarchique du travail de catégorisation

Le travail de catégorisation implique deux activités très différentes en apparence mais difficilement dissociables dans les faits : schématiquement, il recoupe d'une part l'élaboration théorique de types de catégories – travail assigné aux juristes, et en particulier à des acteur·trices occupant une position élevée dans la pyramide hiérarchique de la Cour de Cassation –, et d'autre part l'attribution de labels à certains groupes de mots présents dans les décisions correspondant à ces catégories prédéfinies – travail attribué au moteur d'apprentissage automatique et à l'équipe d'annotation. S'entremêlent ainsi dans la construction du système de catégorisation de la pseudonymisation les modes « aristotélicien » et « prototypique » de classification identifiés par Susan L. Star et Geoffrey Bowker (Bowker & Star, 1999, p.62). Le premier désigne un processus théorique de détermination d'un cadre de classification binaire au sein duquel les entités à classifier sont organisées de façon univoque. Le second correspond à une démarche empirique, par laquelle les catégories émergent du rapprochement analogique de certains objets.

L'élaboration théorique des catégories et la définition des types de contentieux sur lesquels les entités leur correspondant ont vocation à être pseudonymisées est une tâche conceptuelle reposant sur l'expertise juridique. Pour répondre à cet impératif, les juristes du pôle *open data* travaillent en étroite collaboration avec des magistrat·es issus des chambres de la Cour de Cassation et de différentes cours d'appel, réunis dans des groupes de travail. Sur la base de leur connaissance approfondie de leurs contentieux de spécialité et d'échanges collectifs autour de propositions établies par les profils juridiques du pôle *open data*, ces groupes de travail rendent des rapports déterminant les grandes règles de la pseudonymisation. Ces documents précisent ainsi les catégories d'éléments ayant vocation à être occultés en fonction des types de contentieux, ainsi que les termes par lesquels ceux-ci doivent être remplacés (il s'agit au stade de la rédaction de cet article de documents confidentiels). Le pôle *open data* est soumis à une obligation de respect des principes édictés dans ces rapports, auxquels il est fréquemment fait référence comme à une « bible » lors des réunions d'équipe.

En théorie, les opérations de définition des catégories pourraient se limiter au travail conceptuel des groupes de travail – fixé au sein de ces rapports –, l'équipe technique ainsi que les annotateurs étant supposés appliquer à la lettre ces orientations. Le travail de catégorisation concentre pourtant également une part importante de l'activité des membres du pôle *open data*. L'exécution stricte des consignes de catégorisation ne fonctionne pas complètement en pratique, soit parce que la matérialisation technique de ces règles (c'est-à-dire, en particulier, leur traduction algorithmique) nécessite leur fléchissement, soit parce que les cas particuliers auxquels les membres du pôle *open data* sont confrontés au quotidien ne correspondent pas pleinement aux catégories définies en amont. Le travail de l'équipe d'annotation, consistant à attribuer des catégories aux entités réelles présentes dans les décisions à pseudonymiser,

et supposé en ce sens représenter une simple exécution de consignes établies, comporte en réalité une part importante de conceptualisation catégorielle – cette marge d'autonomie faisant écho aux stratégies de négociation des contraintes mises en œuvre par les « *street level bureaucrats* » étudiés par Vincent Dubois (Dubois, 2015). Confronté·es aux décisions bien réelles, et aux informations identifiantes particulières qu'elles contiennent en fonction des cas d'espèce concernés, les agents de l'équipe d'annotation se trouvent fréquemment en prise avec des termes ne correspondant pas exactement aux catégories préétablies. C'est le cas lorsque certaines entités sont susceptibles d'appartenir à plusieurs catégories (voir l'exemplification à suivre avec la catégorie « établissement ») entre lesquelles les agents sont forcés de choisir, contribuant de fait à en déterminer les contours, ou lorsque certaines entités identifiantes de façon évidente ne correspondent à aucune catégorie. Dans ce dernier cas, l'expérience empirique des agents d'annotation conduit parfois à faire évoluer le cadre catégoriel, et notamment à ajouter à l'édifice conceptuel de la pseudonymisation de nouvelles catégories. L'équipe d'annotation, en contact direct avec la matière des décisions, donne par l'exercice de ses activités une dimension prototypique au système de classification de la pseudonymisation.

La parole « sacrée » des groupes de travail ne pouvant plus être modifiée après l'écriture des rapports pour adapter les catégories définies aux évolutions de l'expérience empirique, le pôle *open data* s'est doté d'un instrument au statut intermédiaire ayant vocation à lier les logiques déductives et inductives à l'œuvre dans la définition des catégories. Il s'agit d'un « guide d'annotation », document charnière issu de la rencontre des préconisations des groupes de travail, des problèmes empiriques rencontrés par les agents de l'équipe d'annotation et de la concertation des profils techniques et juridiques du pôle *open data*. Document d'une quinzaine de pages, le guide d'annotation contient la liste des catégories conceptuelles sous-tendant l'annotation, des exemples concrets d'entités (dont la liste est quasi-exhaustive pour certaines catégories) ainsi que des instructions pour le traitement de certains cas particuliers. Il est mis à jour régulièrement en fonction des nouveaux cas d'espèce auxquels l'équipe se trouve confrontée. Ces évolutions font l'objet de longues discussions entre la cheffe de l'équipe d'annotation (faisant remonter les difficultés, incompréhensions et éventuelles incohérences rencontrées par ses agents lors du traitement des décisions), les technicien·nes (garantnes de la faisabilité technique de mise en œuvre des changements demandés) et les juristes (responsables d'un respect minimal des orientations données par les groupes de travail). Fruit de ces négociations, le guide d'annotation fait l'objet de démarches d'appropriation et de réécriture par les différents acteurs, qui en développent une compréhension et un usage propre tout en garantissant un certain niveau de cohésion entre les membres de l'équipe ; il peut en cela être qualifié d' « objet-frontière » de la pseudonymisation (Star et Griesemer, 1989), opérant dans l'épaisseur des versions successives la jonction entre les dynamiques déductives et inductives de structuration des catégories.

Loin de constituer une exception ou une erreur temporaire liée à la phase de construction de l'outil d'IA, la réalisation d'opérations de réflexion et de création catégorielle par les annotateur·trices constitue une part structurelle de ce type

de travail. C'est pour ces l'exécution de ces tâches conceptuelles qu'une part de travail humain demeure indispensable au fonctionnement de l'IA – la machine algorithmique étant elle-même particulièrement performante pour l'exécution systématique de règles, mais incapable de se distancier des instructions catégorielles fournies en entrée.

Distribution du travail de catégorisation : l'exemple de la catégorie « établissement »

L'émergence de la catégorie « établissement »

On l'a vu, logiques déductives et logiques inductives s'enchevêtrent et se superposent dans la définition des catégories sous-tendant la pseudonymisation. L'exemple du cycle de vie d'une catégorie particulière permet d'illustrer cette idée pour mieux la comprendre.
À côté des catégories classiques (nom, prénom, date de naissance, adresse), le pôle *open data* fonctionne avec d'autres catégories recouvrant des réalités parfois plus difficilement identifiables. C'est le cas de la catégorie « établissement », correspondant à des lieux accueillant du public (écoles, hôpitaux, musées, églises, hôtels…) au sein desquels des événements mentionnés dans les décisions auraient pu se dérouler, et dont la citation non occultée dans les décisions pourrait permettre d'identifier de façon indirecte les personnes impliquées.

Ajoutée tardivement à l'édifice de la pseudonymisation, cette catégorie met en tension l'expertise des acteur·trices concerné·es. Non prévue en tant que telle par les rapports des groupes de travail, elle émerge en effet progressivement de la confrontation des règles conceptuelles à l'expérience empirique de codage. Les trois extraits suivants, issus de réunions et séances de travail du pôle *open data*, permettent de suivre le processus hybride mêlant empirie et théorie de création d'une catégorie.

[Extrait 1 : réunion d'équipe janvier 2021, notes de terrain]
Émergence de la catégorie « établissement » de la catégorie « personne morale »
Préalablement à la communication des nouvelles catégories édictées par le groupe de travail à l'équipe d'annotation, les technicien·nes travaillent à leur intégration sur l'interface d'annotation. À cette occasion, l'équipe EIG analyse des exemples de décisions et se pose la question de la catégorisation de types d'entités spécifiques.

EIG :	Est-ce qu'on annote les personnes morales : les écoles, les hôpitaux, les casernes… ? S'il y a une histoire de viol à l'ARS [*Agence Régionale de Santé*], on saura ce que c'est ?
Cheffe d'équipe :	Il faut dissocier le cas où l'institution est partie et où c'est un lieu. Le moteur ne saura pas faire la différence, à part lire la décision et voir le contexte… En tous cas c'est sûr que les AAI [*Autorités Administratives Indépendantes*], il faut pas les annoter [*il s'agit d'administrations publiques*]

uniques qui doivent rester « en clair » dans les décisions au titre de la reddition de comptes]

EIG : Est-ce que la dichotomie, la frontière est claire ?

Cheffe annotation : Je pense pas que ce soit clair pour les agents. Mais c'est difficile de faire une liste des cas particuliers AAI, instances de droit du travail, préjudice corporel…

EIG : On peut créer une nouvelle catégorie « institution / établissement » [*pour les entités différentes des AAI et assimilés*] ?

Cheffe d'équipe : Attention, tout ça diverge des recommandations du groupe de travail qu'on vient de remettre

EIG : Oui mais bon, si le travail n'a pas été bien fait…

Cheffe d'équipe : Est-ce qu'on peut cocher « personne morale » dans NOMOS [*logiciel accessible aux magistrats ayant participé aux groupes de travail*] et en interne à la Cour de Cassation créer une distinction pour notre traitement ? C'est possible au niveau technique ?

[Extrait 2 : réunion d'équipe février 2021, notes de terrain]
Tentative de définition des contours de la catégorie
Suite à la précédente réunion, la catégorie « établissement » est mise en place sur l'interface d'annotation. Ses contours, paraissant originellement clairs, résistent toutefois à la confrontation à de nouveaux cas d'espèce. Après avoir individuellement tenté d'établir des listes de lieux correspondant à la catégorie, l'équipe tente d'en construire une définition.

Data scientist : Non mais là « maison » c'est trop loin, il faut l'enlever

EIG : Ouais mais va dire à un agent que « foyer » il faut annoter mais pas « maison »…

Cheffe annotation : Et sur « auto-école » aussi, moi je trouve qu'on va trop loin

Juriste : Sinon on met « boucherie », « fromager »… faut pas perdre de vue le but, les établissements c'est les structures d'accueil. Mais cette catégorie j'ai du mal à la situer…

Data scientist : On a du mal à tout définir, mais pour le nouveau modèle [*d'apprentissage automatique*] il faut qu'on ait des règles vraiment précises (…). Moi il faut juste qu'on m'énonce une règle qui soit définie et finale et après c'est pas compliqué à appliquer

EIG : Il faut faire un croisement entre l'expérience qu'on a de ta part…

Data scientist : Mon expérience c'est très empirique, il faut voir avec l'expérience de l'équipe d'annotation et l'expérience que [les juristes] ont du GT [*Groupe de Travail*], il faut arbitrer et après on fera évoluer

[Extrait 3 : Entretien et observation du travail d'un agent d'annotation, avril 2021]
La catégorisation dans l'annotation

Le travail réalisé dans les réunions d'équipe, auxquelles les agents de l'équipe d'annotation ne participent pas, aboutit à la mise à jour du guide d'annotation qu'iels sont supposés suivre dans leur travail quotidien. L'équipe d'annotation s'est donc vu présenter cette nouvelle catégorie « établissement », à différencier désormais de la catégorie générale « personne morale » employée précédemment. Ce changement ne se fait pas sans difficultés, à la fois sur le principe et dans les faits, contribuant à donner un corps particulier à cette catégorie. L'agente avec qui je passe la journée annote des décisions en ma présence tout en commentant les évolutions récentes de l'interface et des catégories d'annotation.

Annotatrice : Les jeunes [*l'équipe technique*], ils ont voulu imprimer leur patte, mais je ne suis pas sûre que ce soit vraiment utile. Ces nouvelles catégories (…) comme « établissement », je ne sais pas… « Personne morale », ça me semblait largement suffisant.

[Elle annote une décision, dans laquelle apparaît le terme « camping des forts de Loire »[3]]

Là par exemple, je ne sais pas quoi mettre. Est-ce que c'est un établissement ? Une localité ? Une personne morale ? Il y a un problème avec ces catégories, moi je trouve qu'il y a des choses qui ne sont pas tellement logiques. En réfléchissant au problème, ils ont pas vu tous les cas de figure.

Sociologue : Comment vous choisissez entre ces nouvelles catégories alors ?

[L'annotatrice consulte rapidement le guide d'annotation qu'elle a imprimé et disposé à côté d'elle, et n'y retrouve pas le terme « camping »]

Annotatrice : Il y a des cas particuliers qui n'apparaissent pas dans le guide. Pour ce genre de choses j'avais fait toute une liste, j'en avais fait 20 ou 30, j'avais dit : « faites-nous une liste », parce qu'on a plein de cas de figure différents, est-ce qu'on doit le mettre, pas le mettre… (…) Moi je pense aux personnes qui sont dans la décision, est-ce que ça peut permettre de les identifier, est-ce que ça peut leur poser des problèmes ?

Sociologue : Donc là c'est vraiment au cas par cas, y a pas une règle générale qui a été édictée et que vous suivez ?

Annotatrice : Oui voilà, moi je signale tous ces cas-là, comment on fait ? (…) Moi je sais que je sur-annote beaucoup, je vais même jusqu'à anonymiser le lieu des prisons, je le fais parce que c'est considéré entre guillemets comme un domicile, donc je le fais, mais je sais que pas tout le monde ne le fait (…). On est tous différents par rapport à ça.

[Dans le doute, l'annotatrice termine de traiter la décision, annote l'ensemble de l'expression « camping des forts de Loire » en tant qu'« établissement » et note sur un document annexe ce cas particulier qu'elle compte signaler à sa cheffe d'équipe]

[3] L'expression est anonymisée dans le présent article

Suite à plusieurs retours de l'équipe d'annotation faisant part de ce type de doutes concernant la catégorie « établissement » lors de la phase de test de l'interface d'annotation, le guide a été étoffé. À cette catégorie a été associée une liste de plus d'une cinquantaine d'exemples, se voulant un répertoire quasi-exhaustif des mots à labelliser comme tel. Ces cas particuliers sont issus à la fois de réflexion *in abstracto* de la part de l'équipe technique et juridique et d'exemples remontés par l'équipe d'annotation. Le flou conceptuel entourant la catégorie (le guide la définit ainsi : « La catégorie établissement regroupe principalement les lieux d'accueil du public ») est ainsi compensé par le recours à une logique empirique. Dans d'autres cas, la confrontation de l'équipe d'annotation à l'imprécision conceptuelle de certaines catégories – à laquelle iels peinaient à donner corps – a conduit après de nombreux échanges à la suppression de celles-ci.

On peut tirer plusieurs fils d'analyse de ces fragments du travail de l'équipe *open data*.

Enchevêtrement des logiques conceptuelles et empiriques

Ces extraits illustrent les allers retours permanents entre logiques conceptuelles et empiriques qui conduisent à l'élaboration des catégories d'entités pseudonymisées. L'ajout de la catégorie « établissement » est issu de la confrontation des recommandations théoriques des groupes de travail aux situations empiriquement constatées dans les décisions, faisant apparaitre un manque dans l'architecture conceptuelle de la pseudonymisation. Le constat empiriquement formé de la possibilité de réidentifier des individus à partir du nom de certains lieux (par exemple lorsqu'un « viol » est commis dans une structure unique recevant du public) conduit à faire évoluer le système de classification. Le processus de construction de la nouvelle catégorie « établissement » voit s'enchevêtrer en permanence recherche de cohérence conceptuelle dans la définition de la catégorie et tâtonnements empiriques (visibles dans l'extrait 2) pour en déterminer progressivement les contours à partir d'exemples et d'analogies. L'un des magistrats présents au début du chantier de pseudonymisation par IA souligne en entretien la nécessaire articulation de ces deux logiques dans la construction du système de catégorisation :

Magistrat :	Une fois qu'on a eu la cellule d'anonymisation, concrètement, c'étaient des échanges avec l'équipe, des remontées d'informations pour dire « là j'ai eu telle décision, y avait ça, est-ce que c'est pas problématique ? ». Donc c'est ça, les allers retours entre les objectifs qu'on connaît et la pratique qu'on découvre au fur et à mesure (…) C'était beaucoup de tâtonnement…
Sociologue :	C'était plutôt empirique comme démarche, donc ?
Magistrat :	Totalement empirique oui. Je dirais qu'au départ ça ne peut qu'être empirique, et ça ne peut que l'être de manière constante en fait, parce que ces risques [*de réidentification*] doivent être constamment évalués à

> mesure de l'avancée de la diffusion et avec des nouvelles techniques qui vont apparaître. Donc peut être qu'à un moment ça va être fixé dans un texte, mais il y a un équilibre à trouver (…). En fait, c'est vraiment cette réflexion qui associe l'état de la technique, l'évaluation des risques, à la fois dans l'observation, et de trouver le mieux en fait…

Positionnement ambivalent de l'équipe d'annotation

Cet entremêlement permanent des logiques conceptuelles et empiriques place l'équipe d'annotation dans une position particulière pour la mise en œuvre de la pseudonymisation des décisions de justice, dont l'importance ne correspond pas forcément à sa place statutaire dans la pyramide hiérarchique du projet.
Situés en bas de l'échelle hiérarchique, les annotateur·trices ne disposent pas d'informations précises sur la façon dont les décisions affectant directement leur travail sont prises (concernant les catégories, leur contenu ou le format de l'interface) – ce qui les conduit parfois, à l'instar de l'annotatrice présentée dans l'extrait 3, et conformément à des positions observées dans les travaux consacrés au « travail du clic », à exprimer une certaine méfiance par rapport aux choix effectués en amont (Roberts, 2020).

Pourtant, dans les faits, les activités et réflexions de l'équipe d'annotation revêtent une importance centrale pour la conception et la matérialisation des catégories. D'une part, durant la phase de rodage du système de pseudonymisation, les retours de l'équipe d'annotation sont essentiels pour l'affinage et, éventuellement, la reconfiguration des catégories définies théoriquement. Si d'autres membres de l'équipe – le designeur et les *data scientists* – travaillent également sur des jeux de données leur permettant d'identifier certains cas particuliers potentiellement problématiques, l'équipe d'annotation est la seule à disposer d'une vision exhaustive des décisions traitées. Les cas particuliers considérés comme problématiques (le « camping » dans l'extrait cité) sont remontés – de façon inégale en fonction des agents – en temps réel et lors de quelques réunions dédiées.
D'autre part, une fois le modèle mis en service, le travail de l'équipe d'annotation contribue à donner corps aux catégories prévues. Les agents d'annotation bénéficient d'une certaine liberté d'appréciation pour traiter les situations inédites en prenant en compte les règles du guide d'annotation, le contexte des décisions, les situations particulières qui les caractérisent et les conséquences de leur publication pour les personnes concernées (qui occupent toujours une place importante dans leur travail) – c'est pourquoi leur mission essentielle ne peut être totalement automatisée. En réalisant des micro-choix d'attribution de catégories à certaines entités, sans que ceux-ci soient toujours conformes aux instructions du guide de pseudonymisation (par inattention, par mécompréhension ou par choix) – ni complètement homogène au sein de l'équipe, comme le souligne l'extrait 3 – les annotateur·trices font exister concrètement des catégories dont la définition dans les rapports de groupes de travail ou le guide d'annotation demeure conceptuelle. En y associant une par une des centaines d'expressions particulières, iels donnent une forme tangible à ces entités, qui est susceptible de ne pas tout à fait

correspondre à ses contours théoriques. Largement invisibilisé, ce travail de catégorisation réalisé par une équipe située en bas de la pyramide hiérarchique et considérée comme exécutante occupe pourtant une place centrale dans la définition, à la fois conceptuelle et concrète, des catégories sous-tendant la pseudonymisation et de la façon dont celle-ci prend finalement forme.

Nécessaire clarté conceptuelle et absence d'autonomie du moteur d'IA

Finalement, les extraits présentés, et en particulier l'extrait 2, permettent de réintroduire la dimension machinique associée à ces tâches de catégorisation. S'il est indispensable d'aboutir *in fine* à des catégories conceptuellement claires correspondant à un ensemble déterminé d'éléments, c'est pour mieux guider le travail de l'équipe d'annotation mais surtout pour rendre possible la construction algorithmique du moteur d'IA. La machine n'est en effet apte qu'à reconnaître des entités définies, sur la base d'exemples annotés de façon cohérente et homogène. L'ensemble des processus théoriques et empiriques de construction des catégories, d'enquête et d'adaptation de l'annotation aux situations particulières, de négociation avec les règles édictées par les groupes de travail façonnent directement le moteur d'apprentissage automatique et les résultats que celui-ci fournit.
Ces impératifs techniques peuvent également constituer une ressource pour le pôle *open data* face aux contraintes auxquels il est soumis : comme le met en évidence la fin de l'extrait 1, s'il n'est pas envisageable de s'opposer frontalement aux recommandations des groupes de travail, les spécificités du travail technique peuvent servir à la fois d'alibi et de voile à des évolutions de fond (ici la création d'une nouvelle catégorie) tout en ménageant l'édifice conceptuel et institutionnel de la pseudonymisation.

Conclusion

Le cas étudié de la conception d'un algorithme de pseudonymisation automatique des décisions au sein de la Cour de Cassation fait apparaître l'imbrication du développement de l'IA dans des agencements organisationnels, institutionnels et professionnels qui contribuent à en cadrer les orientations et le fonctionnement. Émerge par le biais de l'enquête ethnographique l'image d'une technique sous-tendue par une « infrastructure informationnelle » multiple (Bowker et Star, 1998 ; Denis, 2018). Instrument de traitement automatisé de l'information, l'IA repose sur la conjonction de données entrantes (les décisions de justice intègres), d'un système de classification évolutif (catégories et règles de codage), d'ensembles d'annotations et de corrections, ainsi que de modèles mathématiques dont les paramètres sont régulièrement optimisés.

Chacune de ces composantes est associée aux activités d'une équipe hétérogène d'une vingtaine de personnes, aux expériences, domaines de compétence et statuts professionnels pluriels. Équipe technique (*data scientists*, développeurs et designeur), petites mains annotatrices et responsables juridiques concourent ensemble à la conception et au bon fonctionnement du moteur d'apprentissage « automatique », dont l'autonomie apparaît alors toute relative. Ce n'est en effet qu'à l'issue de séries d'opérations de cadrage, de définition, de simplification, de négociation et de traduction menées autour de l'outil et de ses modalités opératoires que celui-ci est susceptible de devenir opérationnel. Ces épreuves et leur résolution reposent sur un important travail d'articulation réalisé par les parties prenantes au projet, ainsi que sur des configurations organisationnelles et des systèmes de représentation socialement situés – en témoigne la construction de la catégorie « établissement » - qui contribuent à faire de l'IA un outil technique profondément social. Indispensable au traitement du volume de 3 millions de décisions annuelles produites par le système judiciaire, le recours à outil d'IA ne suppose donc pas la suppression du travail décisionnel humain, mais le déplace en amont de l'acte de pseudonymisation en tant que tel, au moment de la détermination du système de classification guidant le fonctionnement de la machine.

Dans le cas de la Cour de Cassation, la conception d'un tel outil s'inscrit dans le cadre d'une dynamique plus large de « modernisation » de l'action publique par le recours à des outils numériques (Alauzen, 2019a ; Bezes, 2009). Le soutien de l'écosystème réformateur (DITP, Etalab) à ce projet de pseudonymisation automatique a joué un rôle important dans son cadrage organisationnel et opérationnel. Le recours au programme EIG a notamment conditionné la composition particulière de l'équipe technique, ainsi que sa relative autonomie garantie statutairement.

En ancrant l'analyse dans l'observation d'une équipe conceptrice d'un outil d'IA, cet article s'est intéressé aux liens qui unissent les structures professionnelles et organisationnelles à la définition des objectifs associés à cet instrument, ainsi qu'à l'effet des opérations collectives de cadrage et de catégorisation sur son fonctionnement. Le moment de la conception et les controverses qui y sont associées permettent d'accéder à ce qui, une fois la mise en service de l'outil, deviendra rapidement une « boite noire » (Latour, 1995) aux mécanismes difficilement questionnables. Ce travail ouvre la voie à l'étude d'autres dimensions sociales de la conception et du fonctionnement des IA – en l'occurrence juridiques – notamment s'agissant des spécificités du travail de codage algorithmique, des méthodes de *management* des travailleurs associées à ce type d'outils ou encore d'évolution des positionnements institutionnels induits par la mise en œuvre de projets de cette nature.

Bibliographie

ALAUZEN Marie (2019a). « Plis et replis de l'État plateforme. Enquête sur la modernisation des services publics en France », Thèse sciences, technologies et sociétés, Mines Paris-Tech.

ALAUZEN Marie (2019b). « L'Etat Plateforme et l'Identification Numérique des Usagers. Le processus de conception de FranceConnect », *Réseaux*, 1 (213), pp. 211-239.

BARRAUD DE LAGERIE Pauline, SIGALO SANTOS Luc (2018). « Et pour quelques euros de plus », *Réseaux*, 5 (212), pp. 51-84.

BEZES Philippe (2009). *Réinventer l'État. Les réformes de l'administration française (1962-2008)*, Paris, Presses Universitaires de France.

BOLTANSKI Luc, CHIAPELLO Eve (1999). *Le nouvel esprit du capitalisme*, Paris, Gallimard.

BOWKER Geoffrey C., STAR Susan Leigh (1998). « Building Information Infrastructures for Social Worlds — The Role of Classifications and Standards », in ISHIDA Toru (dir.), *Community Computing and Support Systems*, Berlin, Springer pp. 231-248.

BOWKER Geoffrey C., STAR Susan Leigh (1999). *Sorting things out: classification and its consequences*, Cambridge, MIT Press.

BOYD Danah (2016). « Undoing the Neutrality of Big Data », *Florida Law Review*, 67, pp. 226-232.

CALLON Michel (1986). « Éléments pour une sociologie de la traduction. La domestication des coquilles Saint-Jacques dans la Baie de Saint-Brieuc », *L'Année sociologique*, 36, p. 169-208.

CARDON Dominique (2015). *A quoi rêvent les algorithmes ? Nos vies à l'heure des big data*, Paris, Le Seuil.

CARDON Dominique, COINTET Jean-Philippe, MAZIERES Antoine (2018). « La revanche des neurones. L'invention des machines inductives et la controverse de l'intelligence artificielle », *Réseaux*, 5 (211), pp. 173-220.

CASILLI Antonio (2019). *En attendant les robots. Enquête sur le travail du clic*, Paris, Le Seuil.

CHRISTIN Angèle, ROSENBLAT Alex, BOYD Danah (2015). « Courts and Predictive Algorithms », *Data and civil rights: a new era of policing and justice*, pp. 1-11.

CRAWFORD Kate (2021). *Atlas of AI. Power, politics and the planetary costs of Artificial Intelligence*, New Haven and London, Yale University Press.

CRAWFORD Kate, JOLER Vladan (2018). « Anatomy of an AI System: The Amazon Echo As An Anatomical Map of Human Labor, Data and Planetary Resources », [en ligne], consulté le 5/11/2021. URL : https://anatomyof.ai/,

DANZIGER Shai, LEVAV Jonathan, AVNAIM-PESSO Liora (2011). « Extraneous factors in judicial decisions », *Proceedings of the National Academy of Sciences*, 17 (108), pp. 1889-6892.

DENIS Jérôme (2018). *Le travail invisible des données. Elements pour une sociologie des infrastructures scripturales*, Paris, Presses des Mines.

DENIS Jérôme, PONTILLE David (2012). « Travailleurs de l'écrit, matières de l'information », *Revue d'anthropologie des connaissances*, 6 (1), pp. 1-20.

DESROSIERES Alain, THEVENOT Laurent (2002). *Les catégories socio-professionnelles*, Paris, La Découverte.

DUBOIS Vincent (2015). *La vie au guichet. Relation administrative et traitement de la misère*, Paris, Seuil.

EDWARDS Paul N., MAYERNIK Matthew S., BATCHELLER Archer L., BOWKER Geoffrey C., BORGMAN Christine L. (2011). « Science friction: Data, metadata, and collaboration », *Social Studies of Science*, 5 (41), pp. 667-690.

ELISH Madeleine C. (2017). « Dont call AI 'Magic' », [en ligne], consulté le 3/11/2018. URL : https://points.datasociety.net/dont-call-ai-magic-142da16db408

ELISH Madeleine C., BOYD Danah (2018). « Situating methods in the magic of Big Data and AI », *Communication Monographs*, 1 (85), pp. 57-80.

EUBANKS Virginia (2018) *Automating inequality : how high-tech tools profile, police, and punish the poor*, New-York, St Martin's Press.

FORSYTHE Diana E (2001). *Studying those who study us. An anthropologist in the world of artificial intelligence*, Stanford, Stanford University Press.

GANASCIA Jean-Gabriel (2017). *Le mythe de la Singularité. Faut-il craindre l'intelligence artificielle ?*, Paris, Seuil.

GARDEY Delphine (2008). *Écrire, calculer, classer Comment une révolution de papier a transformé les sociétés contemporaines (1800-1940)*, Paris, La Découverte.

GOËTA Samuel (2016). *Instaurer des données, instaurer des publics : une enquête sociologique dans les coulisses de l'open data*, Thèse de doctorat, Télécom ParisTech.

GOFFMAN Erving (1973). *La Présentation de soi. La Mise en scène de la vie quotidienne I*, Paris, Les Editions de Minuit.

JATON Florian (2019). « “Pardonnez cette platitude” : de l’intérêt des ethnographies de laboratoire pour l’étude des processus algorithmiques », *Zilsel*, 1 (5), pp. 315-339.

LACOUR Stéphanie, PIANA Daniela (2019). « Faites entrer les algorithmes ! Regards critiques sur la “Justice Prédictive” », *Cités*, 4 (80), pp. 47-60.

LATOUR Bruno (1995). *La science en action. Introduction à la sociologie des sciences*, Paris, Gallimard.

LICOPPE Christian, DUMOULIN Laurence (2019). « Le travail des juges à l’épreuve des algorithmes de traitement de la jurisprudence. Premières analyses d’une expérimentation de « justice prédictive » en France », *Droit et société*, 3 (103), pp. 535-554.

NEFF Gina, STARK David (2004), « Permanently Beta: Responsive Organization in the Internet Era », in HOWARD Philipp N., JONES Steve. *Society Online: The Internet in Context*, SAGE Publications, pp. 173-188.

PASQUALE Franck (2016) *Black box society : the secret algorithms that control money and information*, Cambridge, Harvard University Press.
ROBERTS Sarah T. (2020). *Derrière les écrans. Les nettoyeurs du Web à l’ombre des réseaux sociaux*, Paris, La Découverte.

STAR Susan Leigh, GRIESEMER James R. (1989). « Institutional Ecology, “Translations” and Boundary Objects: Amateurs and Professionals in Berkeley’s Museum of Vertebrate Zoology », *Social Studies of Science*, 3 (19), pp. 387-420.

STAR Susan Leigh, STRAUSS Anselm (1999). « Layers of Silence, Arenas of Voice: The Ecology of Visible and Invisible Work », *Computer Supported Collaborative Work*, 1-2(8), pp. 9-30.

STRAUSS Anselm (1985), « Work and the Division of Labor », *The Sociological Quarterly*, 1(26), pp. 1-19.

VAUCHEZ Antoine, WILLEMEZ Laurent (2007). *La justice face à ses réformateurs (1980-2006)*, Paris, PUF.